Adaptive Kriging Metamodels for Expensive-to-Run Electromagnetic Simulations

Sándor Bilicz∗†, Szabolcs Gyimóthy∗, Marc Lambert† and József Pávó∗

∗Budapest University of Technology and Economics, Egry J. u. 18, H-1521 Budapest, Hungary
† Département de Recherche en Électromagnétisme, Laboratoire des Signaux et Systèmes UMR8506 (CNRS-SUPELEC-Univ Paris-Sud), 91192 Gif-sur-Yvette cedex, France
E-mail: bilicz@evt.bme.hu

Abstract—Metamodeling is getting more and more widespread in the domain of electromagnetics to replace the often computationally expensive, time-consuming numerical simulations. This paper presents a novel method of metamodeling based on the kriging interpolation. The input-output function (realized by a numerical simulator) is interpolated using a set of output observations at certain input values (“samples”). The quality of the yielded metamodel strongly depends on the choice of these samples. An algorithm is presented to select samples sequentially by optimizing a given criterion related to the optimality of the specific metamodel being built. This adaptive strategy is illustrated and compared to a classical approach via numerical examples drawn from electromagnetic nondestructive evaluation. The presented examples demonstrate that very precise metamodels can be achieved in the proposed way.

Index Terms—adaptive sampling, kriging, metamodeling, nondestructive testing

I. INTRODUCTION

The solution of electromagnetic (EM) problems often involves computationally expensive and time-consuming numerical simulations. In certain applications (like device optimization), the simulator needs to be run several times with different input parameters. Moreover, in other applications, such as the solution of inverse problems in nondestructive testing, the time consumption can be more crucial, whereas many simulation runs are needed as well. To overcome the difficulties raised by such “expensive-to-run” numerical simulators, the EM model can be replaced by a cheap surrogate, or metamodel. Metamodeling has become a widely used and extensively studied domain during the last decade (see, e.g., [1]).

Many metamodeling approaches treat the EM simulator as a black-box, controlled by a set of input parameters (typically a set of scalars) which than determines the output data (one or more scalars or even functional data). In most cases, the metamodel is simply the interpolation of the input-output (I/O) function, based on a set of observed input-output pairs, “samples” (computed off-line by the simulator at hand) – which interpolation can be performed practically in real-time. Both the choice of the samples and the applied interpolator strongly influences the precision of the yielded metamodel, obviously. A wide range of algorithms have been proposed to chose samples in a certain sense “optimal” way, and different interpolation methods are being used.

Classical Design-Of-Experiments (DOE, see [2], [3] for an overview) provides numerous criteria concerning the distribution of the input samples, such as the well-known Latin hypercube sampling (LHS) [4] or maximin sampling [5]. However, these classical methods do not directly take into account either the modeled I/O function or the applied interpolator. More advanced approaches are adaptive, which means that the sampling strategy is driven by the specific I/O function to be interpolated, or at least, the features of the interpolator are taken into account. An example for the latter (weaker) criterion is [6], where an optimal input sample pattern is aimed which minimizes the uncertainty of the kriging interpolator to be fitted. Kriging is a stochastic tool for function approximation, discussed in details later on in this paper (see [7] as a recent review or [9] as a specific EM application).

Real adaptive metamodels are commonly achieved by sequential strategies, where the samples are added to the model one-by-one, by optimizing a certain criterion function (involving the already performed observations) in each iteration. A thorough review of sequential sampling is [10], where metamodels based on kriging and radial basis functions (RBF) are concerned, and several criteria of optimality are given.

Mesh-based adaptive sampling strategies (inspired by the meshing methods of the Finite Element Method) are discussed in [11], [12], using nearest neighbor (NN) interpolators to build a metamodel. A meshless adaptive sampling method is proposed in [13].

A recent development of mathematics is the so-called functional kriging, which is an extension of the original theory (of scalar function interpolation) to the interpolation of functional output data [14], [15]. Functional kriging has successfully been applied for the metamodeling of eddy-current nondestructive testing setups, and has been compared to traditional (NN, piecewise linear and RBF) techniques [16]. However, in [16], the samples were chosen in a naive way: a rectangular grid was defined in the input domain, and samples were placed to the grid-nodes.

In the present paper, a functional kriging metamodel is dealt with, but the sampling strategy is adapted both to the modeled I/O function and to the fitted kriging interpolator. The method is a sequential procedure, solving an optimization task
in each cycle in order to reduce the uncertainty of the resulted interpolator as much as possible. Our method is a combination of the one discussed in [17] (based on the so-called jackknife variance estimation) and a slight improvement inspired by an idea occurring in [10].

Let us highlight the difference between the present sampling method and the others performed by optimization algorithms. From the vast domain of methods we mention as an example for the latter only the Efficient Global Optimization (EGO) algorithm: [18] in EM device optimization and [9] in EM inverse problems. EGO is designed to find the global minimum of an objective function, thus, most of the samples are concentrated to the “promising” regions of the input space. On the contrary, the present sampling strategy aims to improve the quality of the supported surrogate model all over the input space.

II. Metamodelling by functional kriging

A. The EM model as a black-box

Let us consider an EM model controlled by \( p \) input parameters, collected into a vector \( \mathbf{x} \). The forward operator (or I/O function) is usually realized by a numerical simulator (thus, often expensive-to-evaluate), and defines the output corresponding to an input \( \mathbf{x} \) as:

\[
y_\mathbf{x}(t) = \mathcal{F}[\mathbf{x}],
\]

being a function of \( t \in T \) (not necessarily time, but also, e.g., position where a measurement is carried out). In the presented method, no further information is needed about the model (i.e., can be seen as a black-box), which makes the approach generally applicable. For the sake of simplicity, let us assume that all input parameters are linearly scaled to the \([0,1]\) unit interval, and all conceivable input vectors are in the \([0,1]^p\) \( p \)-dimensional unit-hypercube, called the input space \( \mathcal{X} \). The Euclidean norm is used to define distance in \( \mathcal{X} \) in our case.

B. Functional kriging

The goal of metamodelling is to find a cheap approximation for the operator \( \mathcal{F} \). A 2-step method is proposed:

1) Choose some input vectors \( \mathbf{x}_i \) and observe the I/O function by computing the corresponding \( y_{\mathbf{x}_i}(t) \) outputs, i.e., generate a sample-set. Then fit a functional kriging interpolator to the samples.

2) Use the kriging interpolator to predict the sought \( y_{\mathbf{x}_i}(t) \) output function at any arbitrary \( \mathbf{x} \in \mathcal{X} \) input vector (not included in the sample-set).

The time-consuming first step is done only once (by experts, equipped with the numerical simulator), whereas the second step (performed by the end-user as many times as a prediction is needed) just relies on the pre-computed sample-set, yielding an off-line method. In this section, let us forget about the question how the samples have been chosen (this will be detailed in Section IV), and now focus on the kriging interpolation only.

Functional kriging models the function \( y_{\mathbf{x}_i}(t) \) by a Gaussian random process \( \hat{y}_{\mathbf{x}_i}(t) \) [14, [15]. Our goal is to predict the process \( \hat{y}_{\mathbf{x}_i}(t) \) by using a set of observations performed at \( n \) input vectors \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; \hat{y}_{\mathbf{x}_1}(t), \hat{y}_{\mathbf{x}_2}(t), \ldots, \hat{y}_{\mathbf{x}_n}(t) \). The predictor process \( \hat{y}_{\mathbf{x}_i}(t) \) is sought in the linear form

\[
\hat{y}_{\mathbf{x}_i}(t) = \sum_{i=1}^{n} \lambda_i(\mathbf{x}_i)\hat{y}_{\mathbf{x}_i}(t).
\]

The coefficients \( \lambda_i(\mathbf{x}) \) are chosen in a certain sense optimal way, namely, the prediction error must satisfy:

\[
\mathbb{E}\left[\hat{y}_{\mathbf{x}_i}(t) - y_{\mathbf{x}_i}(t)\right] = 0, \forall t \in T,
\]

\[
\int_T \mathbb{E}\left[\hat{y}_{\mathbf{x}_i}(t) - y_{\mathbf{x}_i}(t)\right]^2 dt \rightarrow \text{the smallest possible}.
\]

Equation (3) expresses the unbiasedness of the prediction and criterion (4) formalizes the sense of optimality. The predictor (2) satisfying both conditions (3) and (4) is called the Best Linear Unbiased Predictor (BLUP).

The functional random processes \( \hat{y}_{\mathbf{x}_i}(t) \) are assumed to be stationary in \( \mathbf{x} \), i.e., their mean does not depend on \( \mathbf{x} \):

\[
\mathbb{E}[\hat{y}_{\mathbf{x}_i}(t)] = m(t),
\]

furthermore, the covariance is shift-invariant in \( \mathbf{x} \):

\[
\mathbb{E}\left[\hat{y}_{\mathbf{x}_i}(t) - m(t)\right]\left[\hat{y}_{\mathbf{x}_j}(t) - m(t)\right] = k(\mathbf{x}_i - \mathbf{x}_j, t).
\]

In the framework of the so-called ordinary kriging, \( m(t) \) need not be known, the covariance function is enough to compute the optimal coefficients \( \lambda_i(\mathbf{x}) \) in (2) to fulfill both criteria (3) and (4). It can be proven that the optimal coefficients are the solutions of the following linear system of equations:

\[
\begin{pmatrix}
K(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & K(\mathbf{x}_1 - \mathbf{x}_n) & 1 & \lambda_1(\mathbf{x}) \\
\vdots & \ddots & \vdots & \vdots & \vdots \\
K(\mathbf{x}_n - \mathbf{x}_1) & \cdots & K(\mathbf{x}_n - \mathbf{x}_n) & 1 & \lambda_n(\mathbf{x}) \\
1 & \cdots & 1 & 0 & \mu(\mathbf{x})
\end{pmatrix}
= \begin{pmatrix}
K(\mathbf{x} - \mathbf{x}_1) \\
\vdots \\
K(\mathbf{x} - \mathbf{x}_n) \\
1
\end{pmatrix},
\]

where the function \( K(\cdot) \) is the so-called trace-covariance:

\[
K(\mathbf{x} - \mathbf{x}_j) = \int_T k(\mathbf{x}_i - \mathbf{x}_j, t) dt,
\]

and \( \mu(\mathbf{x}) \) is a Lagrange multiplier.

Once the coefficients \( \lambda_i(\mathbf{x}) \) are computed, the prediction for the output function at any unsampled \( \mathbf{x} \) location can be written as:

\[
\hat{y}_{\mathbf{x}_i}(t) = \sum_{i=1}^{n} \lambda_i(\mathbf{x}_i)y_{\mathbf{x}_i}(t).
\]

Obviously, neither the covariance nor the trace-covariance is given directly with the modeled problem, but one has to predict the latter by using the observed data. A popular parameterized covariance model –called Matérn function– has been chosen:

\[
k(\mathbf{x}_i - \mathbf{x}_j) = \sigma_0^2 \left( \frac{2 \sqrt{\nu d}}{\nu} \right)^\nu K_\nu \left( \frac{2 \sqrt{\nu d}}{\nu} \right),
\]

having the parameters \( \sigma_0^2 \) and \( \nu \) controlling the variance and the smoothness of the process, respectively. \( K(\cdot) \) is the gamma-function, \( K_\nu(\cdot) \) is the modified Bessel-function of the second
kind, of order $v$. $d$ is defined as an anisotropic measure of the
distance of inputs $x_i$ and $x_j$:
\[
d = \sqrt{\sum_{k=1}^{p} \left( \frac{x_{i,k} - x_{j,k}}{\rho_k} \right)^2},
\]
(11)
where the coefficients $\rho_k$ are the characteristic correlation
distances of the $k$th parameter. The smaller the $\rho_k$, the smaller
the range of the trace-covariance along the $k$th parameter, i.e.,
the more sensitive the output to this parameter. Thus, the trace-
covariance can be tuned to the possible anisotropic behavior of
the modeled operator $F$. Consequently, one has to predict $2+p$
hyper parameters from the observed data. We use a classical
statistical technique, cross-validation (CV, see, e.g., in [19]) to
fit the Matérn model to the observed samples $(x_i,y_i(t))$. This
prediction is made once before the adaptive sampling starts, by
using samples regularly spaced in the input space. The yielded
optimization problem is solved by the FMINSEARCH Matlab®
routine.

Let us note that theoretically more elegant and possibly
more precise functional kriging models could be achieved by
using directly the covariance function $k(x_i-x_j,t)$ (called the
“total model” in [15]) instead of restricting ourselves to the (8)
trace-covariance. However, then the coefficients $\lambda_i$ would be
functions of both $x$ and $t$, involving much more computational
effort.

III. QUANTIFYING THE PREDICTION UNCERTAINTY: JACKKNIFE
VARIANCE

Our adaptive sampling strategy aims to reduce the un-
certainty of the functional kriging interpolation as much as
possible whenever a new sample is added to the sample-set.
As a Gaussian process is used to model the I/O function, un-
certainty has a straightforward interpretation in this stochastic
framework: the variance of the prediction error can be used to
quantify our confidence in the prediction.

The variance of the prediction error is formally defined as
\[
\sigma^2(x) = \int_T \mathbb{E} \left[ (\hat{\xi}_x(t) - \xi_x(t))^2 \right] dt,
\]
(12)
and can be approximated by $\hat{\sigma}^2(x)$, having a closed-form
expression based on the trace-covariance and the optimal
coefficients $\lambda_i(x)$:
\[
\hat{\sigma}^2(x) = K(0) - \sum_{i=1}^{n} K(x_i-x_i)\lambda_i(x) - \mu(x).
\]
(13)
Though the above formula is widely used (see, e.g., the
algorithm discussed in [9]), some drawbacks are also pointed
out in [17], [20]. Without going deeply into the details (being
out of the scope of the paper), we just note that in [17], instead
of the classical formula (13), the authors propose the so-called
jacknifing method to estimate the variance. Heuristically,
one can easily see that the formula (13) does not depend
directly on the observations $y_i(t), y_{i,2}(t), \ldots, y_{i,n}(t)$, but only
on the covariance model –consequently, a real adaptive sampling
algorithm cannot be based only on (13). On the contrary, the
jackknife variance estimation is computed directly from the
observed data.

To compute the jackknife variance estimation at an arbitrary
input $x$, one has to perform a traditional kriging prediction
according to (9). Then, one of the $n$ samples (say the $j$th one)
has to be left out, and perform a prediction
\[
\hat{y}_{x,j}^{(-j)}(t) = \sum_{i=0}^{n} \lambda_i^{(-j)}(x)y_i(t),
\]
(14)
where the coefficients $\lambda_i^{(-j)}(x)$ are computed according to (7),
but obviously a reduced system (by deleting the $j$th row and
column) has to be solved. By using the prediction $\hat{y}_{x,j}^{(-j)}(t)$, one
computes the $j$th pseudo-value as:
\[
\hat{y}_{x,j}^{(-j)}(t) = n\hat{y}_{x,j}(t) - (n-1)\hat{y}_{x,j}^{(-j)}(t).
\]
(15)
The jackknife variance estimation is computed by using all $n$
pseudo-values:
\[
\sigma^2_{\text{Jack}}(x) = \frac{1}{n(n-1)} \sum_{j=1}^{n} \int_T \left( \hat{y}_{x,j}^{(-j)}(t) - \frac{1}{n} \sum_{k=1}^{n} \hat{y}_{x,k}^{(-k)}(t) \right)^2 dt.
\]
(16)
One possible drawback of jacknifing may be noticed for the
first glance: to compute the variance estimation at a single point $x$, based on $n$ samples, one has to solve one
linear system of $n+1$ equations and of $n$ equations $n$ times.
However, one has to keep in mind that our sampling method
is designed for “expensive-to-run” simulators, i.e., a simulation
run might involve many times higher computational load than
the jackknife computations.

IV. SEQUENTIAL SAMPLING STRATEGY

Our metamodel –the kriging interpolator according to (9)– is
obviously required to be as precise as possible in the sense that
the estimated jackknife variance (16) is as small as possible
all over the input space $X$. This precision can obviously be
influenced by the location of the samples in $X$.

Following the idea in [17], we use a sequential strategy
to select samples in an optimal way. The procedure starts by
a small $n$ number of initial input samples $(x_1,x_2,\ldots,x_n)$,
chosen according to a space-filling design, e.g. LHS or a
factorial design [3]. Observations of $F$ are performed at these
locations (i.e., $y_1(t), y_2(t), \ldots, y_n(t)$ are computed), and
a kriging interpolator (9) is fitted. Then, an iterative loop
starts, inserting new samples one-by-one according to a certain
criterion. As our goal is to reduce the (16) estimated variance,
a straightforward idea would be to search for the global
maximizer of expression (16) in $X$, i.e., the new, $(n+1)$th
input sample would have to satisfy $x_{n+1} = \arg \max_{x \in X} \sigma^2_{\text{Jack}}(x)$. However, in so doing, the new input sample might be inserted
close to an other, already added one [10], [17]. We follow the
approach discussed in [10] and modify the sample insertion
criteria as:
\[
x_{n+1} = \arg \max_{x \in X} \min_{i=1,2,\ldots,n} \|x-x_i\| \sigma^2_{\text{Jack}}(x),
\]
(17)
in order to avoid clustering samples in $X$ by involving the
minimal distance function $\min_{i=1,2,\ldots,n} \|x-x_i\|$. Now the sequential sampling strategy can be defined as:
1) Choose \( n \) initial input samples and evaluate \( F \) there.
2) Search for the next sample \( x_{n+1} \) by solving (17).
3) Compute \( y_{x_{n+1}}(t) = F[x_{n+1}] \).
4) Add \((x_{n+1}, y_{x_{n+1}}(t))\) to the sample-set and increase \( n = n + 1 \).
5) Go to step 2 until a stopping criterion (related to e.g., the maximal \( \sigma_{\text{jack}}(x) \) variance, or simply an upper bound for \( n \)) is met, then the sampling procedure stops.

This procedure heuristically improves the precision of the metamodel being built. However, we highlight that kriging is a global interpolation method, one cannot expect that this explicit sample-insertion strategy always results a maximum variance lower than in the preceding iteration (since improvement at one region might be accompanied by decay in another). Only an overall decreasing tendency can be expected.

To illustrate the concepts above, a simple scalar function and the performance of the adaptive sampling is presented in Fig. 1. Let us notice that the function is relatively easy-to-predict around \( x \approx 0.5 \), since few samples have been inserted here during the 6 iterations. The opposite situation can be observed around \( x \approx 1 \). The maximal jackknife variance is decreasing as the samples are being inserted.

![Function Plot](image1)

(a) The function (line) and its samples (circles).

![Jackknife Variance](image2)

(b) Jackknife variance estimation.

![Objective Function](image3)

(c) Objective function (to be maximized): \( Q(x) = \min_{i=1,2,...} ||x - x_i||^2 \sigma _{\text{jack}}^2(x) \).

Figure 1. An illustrative one-dimensional toy function and the performance of the adaptive sampling (\( f(x) = -579x^4 + 1110x^3 - 684.5x^2 + 141.071x + 2 \), proposed in [17]). Initial situation (with 4 samples) on the left, and after 6 iterations on the right side.

V. Eddy-current testing examples

A. Configurations and EM simulation

To illustrate our metamodeling approach, we present examples drawn from eddy-current testing (ECT), being a widely used technique in nondestructive evaluation. The basic idea of ECT is to induce eddy-currents in the conductive specimen, and to measure the magnetic field of these eddy-currents. Due to enclosed material flaws, the eddy-current distribution and so the magnetic field changes, thus, flaws can be detected. To solve an inverse problem (to characterize a defect based on measured data) in a short time, one must recourse to a fast metamodel, as the numerical simulation of ECT is usually time-consuming.

In our classical ECT configurations, a homogeneous, non-magnetic, conducting \((\sigma = 10^6 \text{ S/m})\) plate is assumed to be affected by cracks. The plate has finite thickness \((d = 1.25 \text{ mm})\) but it is infinite in other dimensions. An air-cored probe coil (JSAEM benchmark coil) –fed with time-harmonic current with a frequency of 150kHz– scans above the damaged zone with a lift-off of 0.5 mm. The variation of the coil impedance due to the crack is measured at \(11 \times 41\) coil locations, spaced regularly on a grid over a rectangular region, i.e., a surface scan is performed. Thus, the output function \( y_k(t) \) is the impedance variation, \( t \) is related to the position of the coil, and \( T \) corresponds to a whole surface scan. Real and imaginary parts of the impedance variations are separately represented in \( y_k(t) \) in order to keep the latter real-valued. All integrals over \( T \) are approximated by sums, as only samples of \( y_k(t) \) are known at \(11 \times 41\) discrete \( t \) values. The cracks are rectangular-shaped, opening to the plate side opposite with the coil (OD type in the ECT terms), and assumed to be infinitesimally thin. The defect configuration is described by 4 geometrical parameters (Fig. 2). Three cases are defined, according to the number of parameters one lets to vary. In Table I, the so-called 2D (2-dimensional input space), 3D and 4D examples are defined by assigning the fixed parameters and the lower and upper bounds for the varying ones.

The model of the EM phenomena is based on the surface integral method; the integral equation is solved numerically by the method of moments. The reciprocity theorem is used to compute the impedance variation (more details in [21]).

![ECT Configuration](image4)

Figure 2. Sketch of an ECT configuration: cross-sections in two planes. The coil scanned area is enclosed material flaws; the eddy-current distribution and associated field changes, thus, flaws can be detected. To solve an inverse problem (to characterize a defect based on measured data) in a short time, one must recourse to a fast metamodel, as the numerical simulation of ECT is usually time-consuming.

In our classical ECT configurations, a homogeneous, non-magnetic, conducting \((\sigma = 10^6 \text{ S/m})\) plate is assumed to be affected by cracks. The plate has finite thickness \((d = 1.25 \text{ mm})\) but it is infinite in other dimensions. An air-cored probe coil (JSAEM benchmark coil) –fed with time-harmonic current with a frequency of 150kHz– scans above the damaged zone with a lift-off of 0.5 mm. The variation of the coil impedance due to the crack is measured at \(11 \times 41\) coil locations, spaced regularly on a grid over a rectangular region, i.e., a surface scan is performed. Thus, the output function \( y_k(t) \) is the impedance variation, \( t \) is related to the position of the coil, and \( T \) corresponds to a whole surface scan. Real and imaginary parts of the impedance variations are separately represented in \( y_k(t) \) in order to keep the latter real-valued. All integrals over \( T \) are approximated by sums, as only samples of \( y_k(t) \) are known at \(11 \times 41\) discrete \( t \) values. The cracks are rectangular-shaped, opening to the plate side opposite with the coil (OD type in the ECT terms), and assumed to be infinitesimally thin. The defect configuration is described by 4 geometrical parameters (Fig. 2). Three cases are defined, according to the number of parameters one lets to vary. In Table I, the so-called 2D (2-dimensional input space), 3D and 4D examples are defined by assigning the fixed parameters and the lower and upper bounds for the varying ones.

The model of the EM phenomena is based on the surface integral method; the integral equation is solved numerically by the method of moments. The reciprocity theorem is used to compute the impedance variation (more details in [21]).
**Table I**

<table>
<thead>
<tr>
<th>Ex.</th>
<th>A (mm) min</th>
<th>A (mm) max</th>
<th>B (mm) min</th>
<th>B (mm) max</th>
<th>L (mm) min</th>
<th>L (mm) max</th>
<th>D (% of d) min</th>
<th>D (% of d) max</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>-1.5</td>
<td>1.5</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>-1.5</td>
<td>1.5</td>
<td>-1.5</td>
<td>1.5</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>90</td>
</tr>
</tbody>
</table>

**B. Practical considerations**

The adaptive sampling strategy starts by some initial samples. In the presented examples, we use a 3-level full-factorial design [3], i.e., assuming a $p$-dimensional hyper-cube as the input space, the initial input samples are placed at the nodes of a rectangular grid having 3 samples in each direction (totally $3^p$ points), see the triangles in Fig. 3. This design is more convenient than e.g., the LHS design: as mentioned in [17], kriging performs badly in the case of extrapolation (i.e., when used for prediction outside the convex hull of the observations obtained so far).

During the sampling process, one has to solve (17) in each cycle, being a highly complicated optimization problem due to the many local maxima (see Fig. 1(c)). We perform an exhaustive search over a pre-defined set of points in $\mathbb{R}^d$—called “candidate points”, following the notation in [17]. However, in [17], only a few candidate points are used (note that we do not solve exactly the same optimization problem). Our point-set for the exhaustive search includes several hundred or thousand points, generated by an LHS design.

The above mentioned bad performance of kriging in the case of extrapolation inspired a slight change in the (16) jackknife variance estimation. To ensure that no extrapolation occurs, the samples at the vertices of the input space are never left out, i.e., the pseudo-values (15) corresponding to such input samples are not computed. Thus, the sum in the expression (16) has only $n - 2^p$ terms in the case of a $p$-dimensional input space. This is why $\sigma^2_{\text{jack}}$ in the toy example (Fig. 1(b)) tends to zero at the bounds of the input interval.

**Table II**

<table>
<thead>
<tr>
<th>Ex.</th>
<th>Initial samples</th>
<th>Candidate points</th>
<th>Test points</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>3 $\times$ 3 grid</td>
<td>900-point LHS</td>
<td>100 $\times$ 100 grid</td>
</tr>
<tr>
<td>3D</td>
<td>3 $\times$ 3 $\times$ 3 grid</td>
<td>2000-point LHS</td>
<td>15 $\times$ 15 $\times$ 15 grid</td>
</tr>
<tr>
<td>4D</td>
<td>3 $\times$ 3 $\times$ 3 $\times$ 3 grid</td>
<td>3000-point LHS</td>
<td>10 $\times$ 10 $\times$ 10 $\times$ 10 grid</td>
</tr>
</tbody>
</table>

To evaluate the performance of a metamodel, the approximate output data obtained from the metamodel are compared to the true output data. In order to be able to draw a general conclusion about the precision, this comparison should be carried out in all points of the input space. As this is numerically intractable, a set of test points is assigned (being uniformly distributed in $\mathbb{R}^d$) where one performs the comparison.

The definition of the initial, candidate and test points are given in Table II for each example. The stopping criterion of our sampling strategy is simply an upper bound for $n$. More sophisticated criteria might be needed for real applications, but this is not our goal here, however.

**C. Results**

The precision of a metamodel is quantified by the interpolation error, defined as

$$
\epsilon(x) = \sqrt{\frac{1}{T} \int_T (\hat{y}_x(t) - y_x(t))^2 \, dt},
$$

(18)

where $\hat{y}_x(t)$ is the output of the metamodel, $y_x(t)$ is the true output, respectively. The error $\epsilon(x)$ is computed in all test points. The maximal and the average (arithmetical mean of all values) error is denoted by $\epsilon_{\text{max}}$ and $\bar{\epsilon}$, respectively. The adaptive sampling is compared to the regular sampling (input samples are at the nodes of a rectangular grid in $\mathbb{R}^d$, as presented in [16]), using the same number of samples. The numerical results are summarized in Table III.

**Table III**

<table>
<thead>
<tr>
<th>Ex.</th>
<th>No. of samples</th>
<th>Maximal error ($\epsilon_{\text{max}}$)</th>
<th>Average error ($\bar{\epsilon}$)</th>
<th>Time (µs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>$4^d$</td>
<td>13.57</td>
<td>7.66</td>
<td>1.94</td>
</tr>
<tr>
<td>5$^d$</td>
<td>11.46</td>
<td>3.14</td>
<td>1.59</td>
<td>1.04</td>
</tr>
<tr>
<td>6$^d$</td>
<td>9.04</td>
<td>3.17</td>
<td>1.01</td>
<td>0.72</td>
</tr>
<tr>
<td>7$^d$</td>
<td>6.96</td>
<td>1.72</td>
<td>0.72</td>
<td>0.29</td>
</tr>
<tr>
<td>3D</td>
<td>$4^d$</td>
<td>14.40</td>
<td>5.11</td>
<td>2.06</td>
</tr>
<tr>
<td>5$^d$</td>
<td>12.86</td>
<td>4.24</td>
<td>1.86</td>
<td>0.70</td>
</tr>
<tr>
<td>6$^d$</td>
<td>10.22</td>
<td>4.46</td>
<td>1.20</td>
<td>0.44</td>
</tr>
<tr>
<td>4D</td>
<td>$4^d$</td>
<td>20.12</td>
<td>10.53</td>
<td>2.30</td>
</tr>
</tbody>
</table>

One can see the extremely low time consumption of the kriging prediction in Table III—which is obvious knowing that the interpolation is nothing but the evaluations of the function $K(\cdot)$ and the solution of (7) (assuming that the covariance parameters have already been estimated at the sampling stage). On the contrary, the precise evaluation of $\mathcal{F}$ (the numerical field computation) involves a computation time in the order of some minutes (depending on the size of the crack, see [21]). Indeed, metamodeling may make possible the real-time solution of e.g., inverse problems of ECT.

The adaptive sampling outperforms the regular one: $\epsilon_{\text{max}}$ in all presented cases, and $\bar{\epsilon}$ almost in all cases (except the first 2D case) is definitely smaller for the adaptive sampling. However, let us note that the decrease of the maximal error is not monotonous when the number of samples increases in the adaptive case. On the contrary, this disadvantage does not occur when one performs regular sampling. The average error decreases monotonously with increasing number of samples in both cases, however.

In Fig. 3, the error $\epsilon(x)$ is plotted over the input space in one of the 2D examples. The definitively better performance of the adaptive metamodel is obvious. The adaptive sampling has inserted much more samples to the domain of deep defects, whereas the shallow ones are represented by only few samples. This inhomogeneous sample distribution is caused by the varying behavior of the forward operator: $\mathcal{F}$ must vary fast in the densely sampled regions, and must be “flat” elsewhere. However, there is no direct relation, as the sample-insertion criteria (17) is not directly based on the distance of output samples (in contrary with the sampling proposed in [13]).
The evaluation of the sample-insertion criterion (17) is getting more and more demanding with increasing number of input dimensions. An alternative of the presented solution (exhaustive search) could be a multistart local optimization, for instance. However, further research would be needed at this point.

VI. CONCLUSIONS

A novel adaptive sampling method for metamodeling has been constructed by combining recent mathematical tools. Samples are sequentially inserted so that the fitted functional kriging interpolator is as precise as possible. The basic idea is the reduction of the jackknife variance of the kriging predictor.

The method tested on classical eddy-current testing configurations. The use of the metamodel appeared to be much faster than the precise forward model. The metamodels yielded by the adaptive sampling are compared to metamodels based on regular (non-adaptive) sampling on the basis of provided precision. The results demonstrate that the new sampling method highly outperforms the traditional method.

Though we have chosen only eddy-current testing examples for the illustration, the approach can obviously be applied in a wider range of black-box modeling problems.

The so-called “curse of dimensionality” rises a bottleneck. The evaluation of the sample-insertion criterion (17) is getting more and more demanding with increasing number of input dimensions. An alternative of the presented solution (exhaustive search) could be a multistart local optimization, for instance. However, further research would be needed at this point.

ACKNOWLEDGEMENTS

This research is partially supported by the Research and Technology Innovation Fund of the Hungarian Government in the frame of the French-Hungarian Bilateral Intergovernmental S&T Cooperation (FR-1/2008) and by DIGITEO cluster’s project.

REFERENCES